Unreliable Inferences about Unobserved Processes

A Critique of Partial Observability Models*

Carlisle Rainey‡ and Robert A. Jackson†

Abstract

Methodologists and econometricians have begun to advocate the partial observability model (aka the split population model) as an approach that enables researchers to draw accurate inferences about effects of explanatory variables on partially observable outcome variables. Although the partial observability model is theoretically-driven and statistically sound, we show that (presumably unavoidable) minor model misspecification can lead to large inferential errors. We use simulations to show that seemingly innocuous model specification errors that have little impact under full observability can lead to large biases, including sign errors, under partial observability.

* We thank Will Moore, Austin Mitchell, participants at the 2012 Southern Political Science Association Annual Conference, and participants at the 2012 Midwest Political Science Association Annual Conference for valuable comments on previous versions of the manuscript. All code and data necessary to replicate the simulations and empirical analysis are available at github.com/carlislerainey/unreliable-inferences.

‡ Carlisle Rainey is Assistant Professor of Political Science, University at Buffalo, SUNY, 520 Park Hall, Buffalo, NY 14260 (rcrainey@buffalo.edu).

† Robert A. Jackson is Professor of Political Science, Florida State University, 531 Bellamy Building, Florida State University, Tallahassee, FL 32306 (rjackson@fsu.edu).

"The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data."

—John Tukey (1986, p. 74)

Introduction

Social scientists often face situations, known as "partial observability," in which two (or perhaps more) distinct processes lead to distinct binary outcomes that can only be observed jointly (Poirier 1980; Abowd and Farber 1982; Nieman 2015).[1] Braumoeller (2003) provides many examples of established literature theorizing such relationships. In some situations, the analyst believes that the key explanatory variable influences only one of the partially observed outcomes (e.g., Braumoeller and Carson 2011), but we focus on the more difficult situation in which the researcher believes that the key explanatory variable affects *both* partially observed outcomes (e.g., Xiang 2010).

Research employing the partial observability model has already weighed in on social scientists' understanding of such important processes and outcomes such as civil wars (Nieman 2015), international conflict and trade (Xiang 2010), IMF agreements (Knight and Santaella 1997, Przeworski and Vreeland 2000, 2002, Vreeland 2003, and Stone 2008), union membership (Abowd and Farber 1982), regulatory compliance (Feinstein 1990, Stafford 2002, Chen et al. 2006, and Wang 2013), network formation (Comola and Fafchamps 2014), credit ratings (Boyes, Hoffman, and Low 1989), agricultural innovation (Dimara and Skuras 2003), health insurance ownership (Amir 2001), and employment discrimination (Heywood and Mohanty 1990, Logan 1996, and Mohanty 2002). Feinstein (1990) suggests the model's usefulness for a wide range of policy studies, and the model appears to hold out promise for investigating numerous other subjects — including

[1] In some applications, researchers observe an event if either unobservable event occurs. In other cases, researchers only observe an event if both occur.

deterrence, treaty compliance, and attitudes and behaviors that are subject to social desirability bias when measured via a survey report (see Beger et al. 2011).

Indeed, recent work suggests that scholars can draw compelling inferences about the effect of the key explanatory variable on each unobserved outcome via the partial observability model (aka the split population model) as long as the researcher can identify at least one other variable that influences only one of the processes (Przeworski and Vreeland 2002; see Nieman 2015 for a strategic variant). However, we present simulation evidence showing that the partial observability model is extremely sensitive to misspecification.

Recent applications are relatively sanguine about employing the partial observability model. However, Meng and Schmidt (1985) counseled economists early on regarding some of the costs of partial observability, arguing that standard errors are much larger when the outcome of interest is only partially observed. They write, "we would not be surprised to find, in a typical application, *t*-ratios to be from two to four times as large under full observability as under partial observability" (Meng and Schmidt 1985, p. 83). Yet standard errors (by design) only reflect the uncertainty due to sampling error. Other sources of error, such as measurement error, missing data, and specification error create additional uncertainty. In the simulations below, we focus specifically on the cost of specification error.

Although innocuous specification errors (particularly of the functional form variety) do not usually introduce much additional uncertainty under full observability, we argue that this is not the case with partial observability models.[2] We argue that an even greater cost of partial observability models (compared to a doubling or even quadrupling of the standard errors as cited above) is the substantial *bias* introduced by small specification errors, which is not reflected in the standard errors.

[2] For example, a functional form error might be simply including x in the model when the actual specification is $\log(x)$.

We use simulations to show that a misspecification that leads to almost no bias under full observability leads to large biases under partial observability. The type of misspecification that we consider is much less serious than including a variable in the incorrect equation (though this seems quite possible in many applications). Instead, we assess a slight misspecification of functional form—a situation in which the logit link function is not quite right. The relationship between the probability of each partially observed event and the explanatory variables is non-linear, but monotonic—specifically, the relationship is quadratic. We view this type of specification error as quite mild for two reasons. First, there is rarely a compelling theoretical rationale for preferring the specific logit functional form over a "non-linear and monotonic" functional form. Second, in logit models of fully observed binary outcomes, this type of misspecification error appears to have almost no effect on the inferences (see our simulations below). Although this mild form of misspecification has little impact on the inferences from full observability logit models, we show that this type of misspecification leads to large errors in the partial observability context—partial observability models are (much) more sensitive to misspecification than their fully-observed counterparts.

The Partial Observability Model

Assume a situation in which a potential dichotomous outcome of interest d_{main} cannot be directly observed. Instead, it can only be observed jointly with another dichotomous outcome $d_{nuisance}$.[3] The researcher only observes the outcome variable y_{obs}, which, depending on the application, might equal one if both $d_{main}$and $d_{nuisance}$ equal one or if either d_{main} or $d_{nuisance}$ equals one. For simplicity, we consider the situation in which y_{obs} equals one if either $d_{main} = 1$ or $d_{nuisance} = 1$. That is, $P(y_{obs}) = P(d_{main}) + P(d_{nuisance}) - P(d_{main}) \times P(d_{nuisance})$. However, all of our findings and suggestions generalize to the other situation as well.[4]

[3] For clarity, we imagine a situation in which one outcome is of interest and the other is largely nuisance. However, our ideas generalize to a situation in which the researcher's interest extends to both outcomes.

[4] In fact, there is an exact mathematical relationship between the two. Specifically, assuming that $y_{obs} = 1$ when

To model the outcome variable y_{obs}, we simply recognize that it is determined jointly by d_{main} and $d_{nuisance}$ and assume that a standard logit model relates the matrix of covariates X to $P(d_{main})$, so that $P(d_{main}) = logit^{-1}(X\beta)$. Similarly, assume that the same model relates the matrix of covariates Z to $P(d_{nuisance})$, so that $P(d_{nuisance}) = logit^{-1}(Z\gamma)$. Then, rewriting the equation for $P(y)$ above, we obtain the partial observability model

$$P(y_{obs}) = logit^{-1}(X\beta) \ + logit^{-1}(Z\gamma) + logit^{-1}(X\beta) \times logit^{-1}(Z\gamma).$$

Using this form, it is straightforward to find the log-likelihood function for β and γ, although the maximization is not trivial.[5]

Simulations

In setting up the simulations, we assume that the researcher does not know the exact functional form relating the explanatory variables to the unobserved outcomes, but make the generous assumptions that she does know that only *x* influences d_{main}, that both *x* and *z* influence $d_{nuisance}$, and that the following relationships hold:

1. $P(d^{main}) = f(x)$, where *f* is a monotonic function
2. $P(d^{nuisance} = 1) = g(x,\ z)$, where g is a monotonic function of both *x* and *z*
3. $P(y) = f(x) = \begin{cases} 1, & d^{main} = 1 \text{ or } d^{nuisance} = 1 \\ 0, & \text{otherwise} \end{cases}$

This is less messy than many applied problems, which almost always involve many more variables, as well as uncertainty about which variables belong in each equation. In

---

$d_{main} = 1$ <u>or</u> $d_{nuisance} = 1$ is equivalent to modeling the probability that $y_{obs} = 0$ when $d_{main} = 0$ <u>and</u> $d_{nuisance} = 0$.

[5] We find that standard "hill-climbing" optimization routines find local maxima—this present a separate, perhaps equally concerning problem to substantive applications. To avoid this, we tried an algorithm suggested by Sekhon and Mebane (1998) that relies on a combination of genetic adaptation and hill-climbing to efficiently locate the maximum (Mebane and Sekhon 2011). This approach consistently located the global maximum in several trial examples, but only for a very large population (i.e., computationally prohibitive) of agents. We found that simply starting a standard hill-climbing algorithm at several sets of random starting values consistently located the global maximum. We use ten sets of starting values in the simulation studies and 200 in the empirical illustration. Additionally, partial observability models are prone to problems with separation (i.e., a coefficient with the value of infinity maximizing the likelihood function). To avoid this problem, we use a weakly informative Cauchy prior with scale parameter 2.5, as suggested by Gelman et al. (2008). In general, this prior tends to improve the inferences from the model.

our hypothetical situation, the researcher knows that the relationships are monotonic and knows exactly which variables belong in each equation, so the misspecification, while always present, is slight. Further, the models are relatively simple—only one variable in the main equation and two variables in the nuisance equation. Indeed, this strikes us as a rather "friendly" scenario for social science research.

We simulate 500 different relationships that fit into our hypothetical scenario and calculate the large-sample properties of the partial observability model. To generate each hypothetical relationship, we do the following:

Assume that the true relationship between $P(d_{main})$ and *x* is given by $P(d_{main}) = \beta_{cons} + \beta_x x + \beta_{x^2} x^2$. Assume that the true relationship between $P(d_{nuisance})$ and *x* and *z* is given by $P(d_{nuisance}) = \gamma_{cons} + \gamma_x x + \gamma_z z + \gamma_{xz} xz + \gamma_{x^2} + \gamma_{z^2} z^2$.

Simulate the parameter vectors β and γ from a uniform distribution that ranges from -1 to 1. Repeat until all of the following conditions are met:

1. Both $P(d_{main})$ and $P(d_{nuisance})$ are bounded between zero and one across the ranges of *x* and *z*.
2. The relationship between $P(d_{main})$ and *x* is monotonic across the range of *x*.
3. The relationship between $P(d_{nuisance})$ and *x* and *z* is monotonic (in both *x* and *z*) across the ranges of *x* and *z*.

These steps ensure that each relationship is monotonically increasing in *x* and *z*, so that the misspecification is relatively mild. Using the procedure above, we calculate the large-sample properties of the estimator using the following steps:

1. Simulate a hypothetical relationship using the steps above.
2. Simulate the explanatory variable *x* by taking 100,000 draws from a uniform distribution from zero to one. Simulate *z* similarly.
3. Using the simulated β from Step 1, calculate $P(d_{main}| x)$. Simulate d_{main} by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(d_{main}| x)$.
4. Similarly, using the simulated γ from Step 1, calculate $P(d_{nuisance}| x, z)$. Simulate $d_{nuisance}$ by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(d_{nuisance}| x, z)$.
5. Calculate $P(y_{obs})$ using the identity $P(y_{obs}) = P(d_{main}) + P(d_{nuisance}) - P(d_{main}) \times P(d_{nuisance})$. Simulate y_{obs} by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(y_{obs})$.

6. Calculate the full observability estimate using the following steps:
    a. Estimate the usual logistic regression model $P(d_{main}) = \Delta^{-1}(b_{cons} + b_x x)$.
    b. Use the large-sample estimates b_{cons} and b_x to calculate the predicted probability that $d_{main} = 1$ when *x* is set to a low value (25th percentile or 0.25) and when *x* is set to a high value (75th percentile or 0.75). Subtract the latter from the former to obtain an estimate of the effect of changing *x* from 0.25 to 0.75 on $P(d_{main} = 1)$. Store this estimate.
7. Calculate the partial observability estimate using the following steps:
    a. Estimate the partial observability logistic regression model

$$P(y_{obs}) = \Delta^{-1}(b_{cons} + b_x x) + \Delta^{-1}(g_{cons} + g_x x + g_z z)\times(1 - \Delta^{-1}(b_{cons} + b_x x)).$$

    b. Use the (large-sample) estimates b_{cons} and b_x to calculate the predicted probability that $d_{main} = 1$ when *x* is set to a low value (25th percentile or 0.25) and when *x* is set to a high value (75th percentile or 0.75). Subtract the latter from the former to obtain an estimate of the effect of changing *x* from 0.25 to 0.75 on $P(d_{main} = 1)$. Store this estimate.
8. Calculate the true effect using the parameters of the hypothetical relationship generated in Step 1 and the true functional form. Store this estimate.

We consider two estimands (potential quantities of interest). First, we examine how well the full and partial observability models estimate the probability of the event of key interest (i.e., $P(d_{main} = 1)$) when the explanatory variable takes on a central value ($x = 0.5$ in our simulations). Second, we estimate how this probability changes as *x* moves from its 25th to 75th percentile (from 0.25 to 0.75 in our example). Although the estimators are equally good in the theoretical sense that they are both consistent but biased in small samples, we show that the partial observability model is quite sensitive to seemingly innocuous specification errors. Indeed, specification errors that have almost no impact on the inferences from a full observability logistic regression model lead to large biases in a partial observability model.

Results

Figure 1 shows how well the full and partial observability models estimate the probability of the event of interest when the explanatory variable takes on a central value and the functional form is slightly misspecified. The figure shows the histogram and density for the large-sample bias in the estimate. The full observability model performs surprisingly well under misspecification. The absolute bias is greater than 0.03 in only about 10% of the simulations. The partial observability model performs surprisingly poorly. The absolute bias is greater than 0.03 in about 85% of the simulations and greater than 0.55 in 10% of the simulations. That is, even with a large data set and a benign scenario for an applied problem, there is a 10% chance that the probability of an event is off by more than half of the possible range! The errors are not symmetrically distributed around the true probability, however. While overestimation is possible, the partial observability model tends to underestimate the occurrence of the event of interest and can drastically underestimate it.

[Insert Figure 1 about here.]

Next, we examine the ability of the partial observability model to estimate how the probability of the event of interest changes as the key explanatory variable increases from a low value to a high value (0.25 to 0.75 in our case). Figure 2 shows the results. Again, notice that the full observability estimate produces much less bias than the partial observability model. The absolute bias in the estimated effect from the full observability model is greater than 0.03 in only 10% of the scenarios. For the partial observability model, the absolute bias is greater than 0.03 in 80% of the scenarios, and greater than 0.25 in 10% of the scenarios. An effect of 0.25 is quite large in the social sciences. This means that the partial observability model is prone to large sample bias — in fact, bias that is larger than many of the effects this model is being used to estimate.

[Insert Figure 2 about here.]

Figure 3 shows the estimated probabilities against the true probabilities for both models when *x* is set at a central value for each of the 500 scenarios. The left panel shows that the full observability estimates are very strongly correlated with the true estimates (Kendall's τ= 0.93)..[6] However, the relationship between the true probability and the

---

[6] Kendall's tau is comparable to Pearson's r, but is a rank-based statistic and thus more robust to data that do not come from a multivariate normal distribution.

estimated probability from the partial observability model is extremely weak (Kendall's τ= 0.20). While seemingly innocuous specification errors have essentially no effect on the bias of the estimator under full observability, the effect can be quite large under partial observability.

[Insert Figure 3 about here.]

Figure 4 shows the scatterplots of the true effect and the estimated effect for both models. As before, the correlation between the true effect and the estimated effect under full observability is extremely strong (Kendall's τ= 0.99). While the partial observability model does a substantially better job of estimating effects than estimating probabilities, the correlation between the true effect and the partial observability estimate is still quite weak (Kendall's τ= 0.62). Only 40 of the 500 (8%) simulations produce a sign error, so the partial observability estimator does tend to get the sign correct. However, the full observability model produces no sign errors.

[Insert Figure 4 about here.]

Conclusion

These simulations demonstrate, what we think is an under-appreciated but critical characteristic of partial observability models—they are quite sensitive to seemingly innocuous specification errors. Researchers commonly interpret standard errors as the uncertainty of the estimate. However, standard errors only capture the uncertainty due to sampling variation. Other sources of are present as well, including specification uncertainty. If a researcher is uncertain about the model specification, then the standard errors from a default model are too small. In the case of full observability, our simulations show that the additional uncertainty is perhaps negligible. But in the case of partial observability, the additional uncertainty is quite meaningful. The biases introduced from specification error of the partial observability model can be quite large, are rarely negligible, and are not captured in the standard errors.

References

Abowd, John M., and Henry S. Farber. 1982. "Job Queues and the Union Status of Workers." *Industrial and Labor Relations Review* 35:354-367.

Amir, Shmueli. 2001. “The Effect of Health on Acute Care Supplemental Insurance Ownership: An Empirical Analysis.” *Health Economics* 10: 341-350.

Boyes, William J., Dennis L. Hoffman, and Stuart A. Low. 1989. “An Econometric Analysis of the Bank Credit Scoring Problem." *Journal of Econometrics* 40: 3-14.

Braumoeller, Bear F. 2003. “Causal Complexity and the Study of Politics.” *Political Analysis* 11:209-233.

Braumoeller, Bear F. and Austin Carson. 2011. “Political Irrelevance, Democracy, and the Limits of Militarized Conflict.” *Journal of Conflict Resolution* 55:292-320.

Chen, Gongmeng, Michael Firth, Daniel N. Gao, and Oliver M. Rui. 2006. “Ownership Structure, Corporate Governance, and Fraud: Evidence from China.” *Journal of Corporate Finance* 12: 424-448.

Comola, Margherita, and Marcel Fafchamps. 2014. “Testing Unilateral and Bilateral Link Formation.” *The Economic Journal* 124: 954-976.

Dimara, Efthalia, and Dimitris Skuras. 2003. “Adoption of Agricultural Innovations as a Two-Stage Partial Observability Process.” *Agricultural Economics* 28(3): 187-196.

Feinstein, Jonathan S. 1990. “Detection Controlled Estimation.” *Journal of Law and Economics* 33:233-276.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. “A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models.” *The Annals of Applied Statistics* 2:1360-1383.

Heywood, John S., and Madhu S. Mohanty. “Race and Employment in the Federal Sector.” *Economics Letters* 33: 179-183.

Knight, Malcolm, and Julio A. Santaella. 1997. “Economic Determinants of IMF Financial Arrangements.” *Journal of Development Economics* 54: 405-436.

Logan, John Allen. 1996. “Opportunity and Choice in Socially Structured Labor Markets.” *American Journal of Sociology* 102:114-160.

Sekhon, Jasjeet S., and Walter R. Mebane. 1998. “Genetic Optimization Using Derivatives.” *Political Analysis* 7(1): 187-210.

Mebane Jr, Walter R., and Jasjeet S. Sekhon. 2011. “Genetic Optimization Using Derivatives: The rgenoud Package for R.” *Journal of Statistical Software* 42: 1-26.

Meng, Chun-Lo, and Peter Schmidt. 1985. “On the Cost of Partial Observability in the Bivariate Probit Model." *International Economic Review* 26:71-85.

Mohanty, Madhu S. 2002 “A Bivariate Probit Approach to the Determination of Employment: A Study of Teen Employment Differentials in Los Angeles County.” *Applied Economics* 34:143-156.

Nieman, Mark David. 2015. “Statistical Analysis of Strategic Interaction with Unobserved Player Actions: Introducing a Strategic Probit with Partial Observability.” *Political Analysis.* Forthcoming.

Poirier, Dale J. 1980. “Partial Observability in Bivariate Probit Models.” *Journal of Econometrics* 12:209-217.

Przeworski, Adam, and James Raymond Vreeland. 2000. “The Effect of IMF Programs on Economic Growth.” *Journal of Development Economics* 62:385-421.

Przeworski, Adam, and James Raymond Vreeland. 2002. “A Statistical Model of Bilateral Cooperation.” *Political Analysis* 10:101-112.

Stafford, Sarah L. 2002. “The Effect of Punishment on Firm Compliance with Hazardous Waste Regulations.” *Journal of Environmental Economics and Management* 44: 290-308.

Stone, Randall W. 2008. “The Scope of IMF Conditionality.” *International Organization* 62:589-620.

Tukey, John W. 1986. “Sunset Salvo.” *The American Statistician* 40:72-76.

Vreeland, James Raymond. 2003. *The IMF and Economic Development*. Cambridge, UK: Cambridge University Press.

Wang, Tracy Yue. 2013. "Corporate Securities Fraud: Insights from a New Empirical Framework." *Journal of Law, Economics, and Organization* 29:535-568.

Xiang, Jun. 2010. "Relevance as a Latent Variable in Dyadic Analysis of Conflict." *Journal of Politics* 72:484-498.

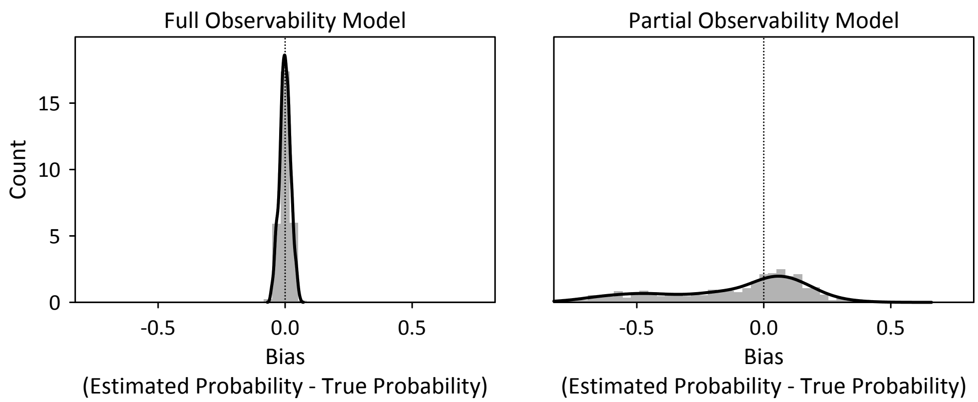

Figure 1: Summarizing the ability of the full and partial observability models to estimate the occurrence probability of the event of interest when the explanatory variable is set to a central value and the functional form is slightly misspecified.

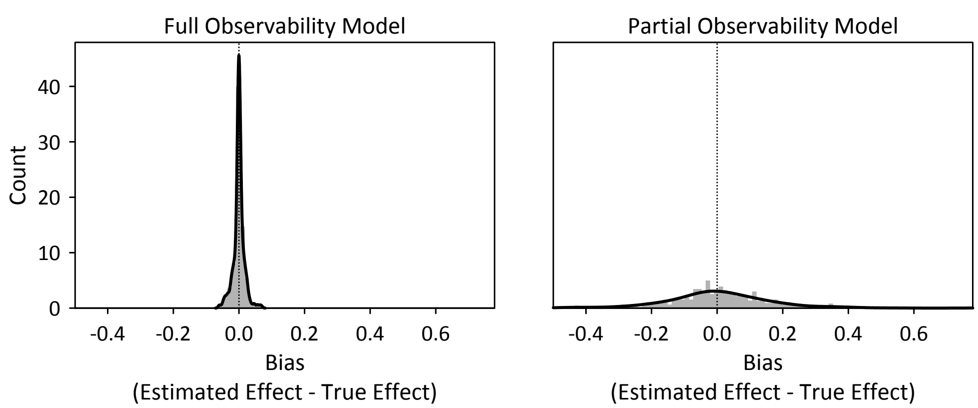

Figure 2: Summarizing the ability of the full and partial observability models to estimate how the occurrence probability of the event of interest changes as the key explanatory variable moves from a low value to a high value when the functional form is slightly misspecified.

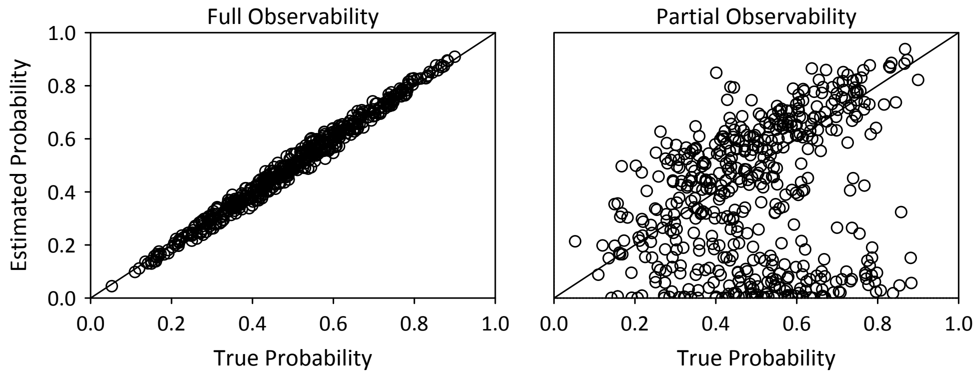

Figure 3: Comparing the relationship between the true probability and the estimated probability of the event of interest for the full and partial observability models when the model is slightly misspecified.

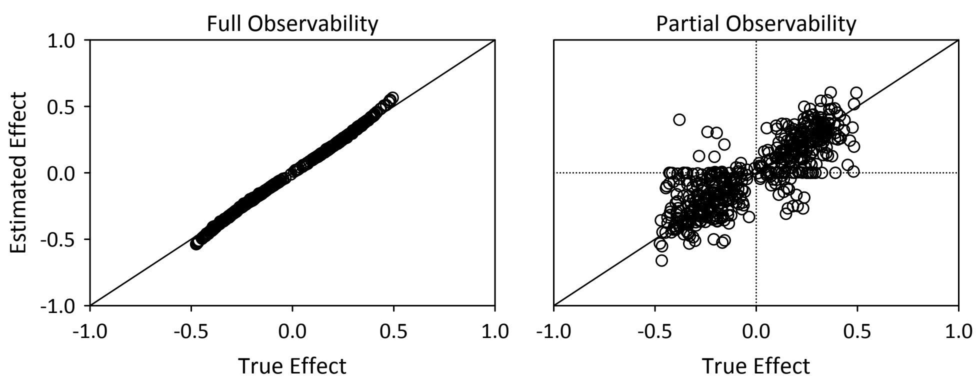

Figure 4: Comparing the relationship between the true effect and the estimated effect on the event of interest for the full and partial observability models when the model is slightly misspecified. The solid 45 degree line is y = x and represents a one-to-one correspondence between the true value and the estimate. If a point falls on the line, then the estimate is exactly correct.